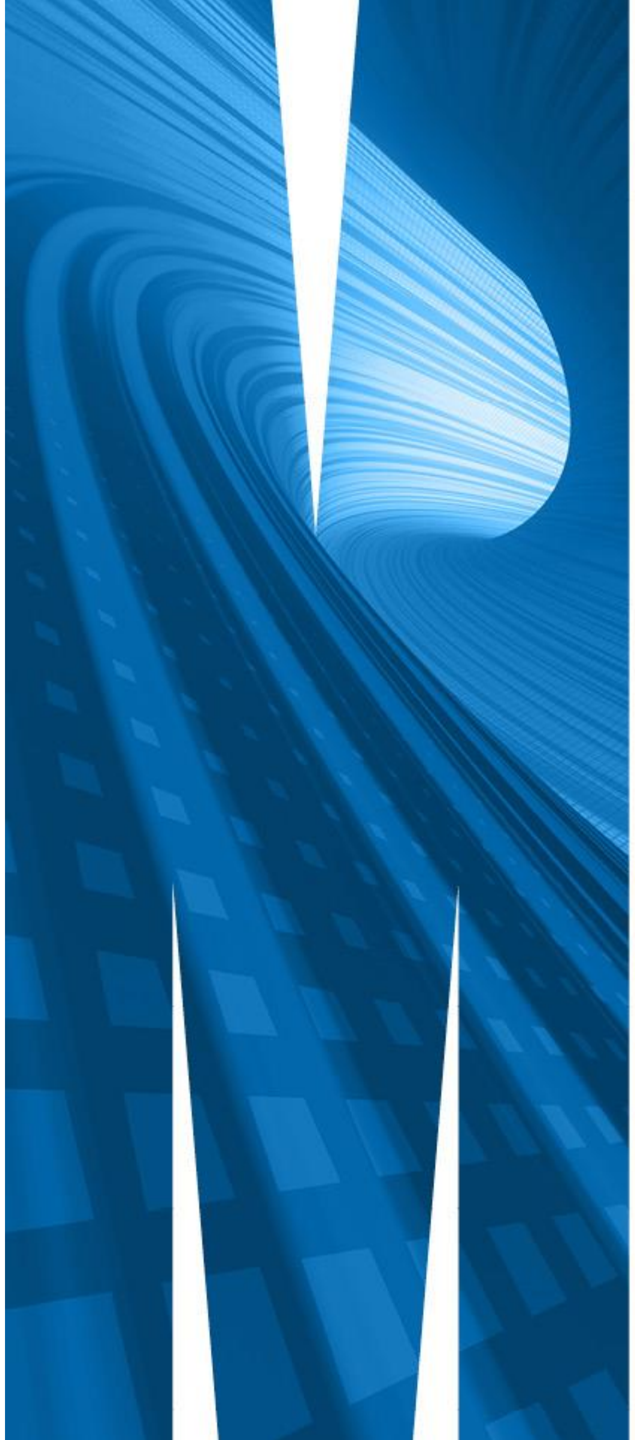
MONASH University

The problem of multiplicity and the use of hierarchical selection rules

Matthew J Page, Monash University, Australia

Cochrane Methods Symposium: Developing robust review protocols with increasingly diverse evidence

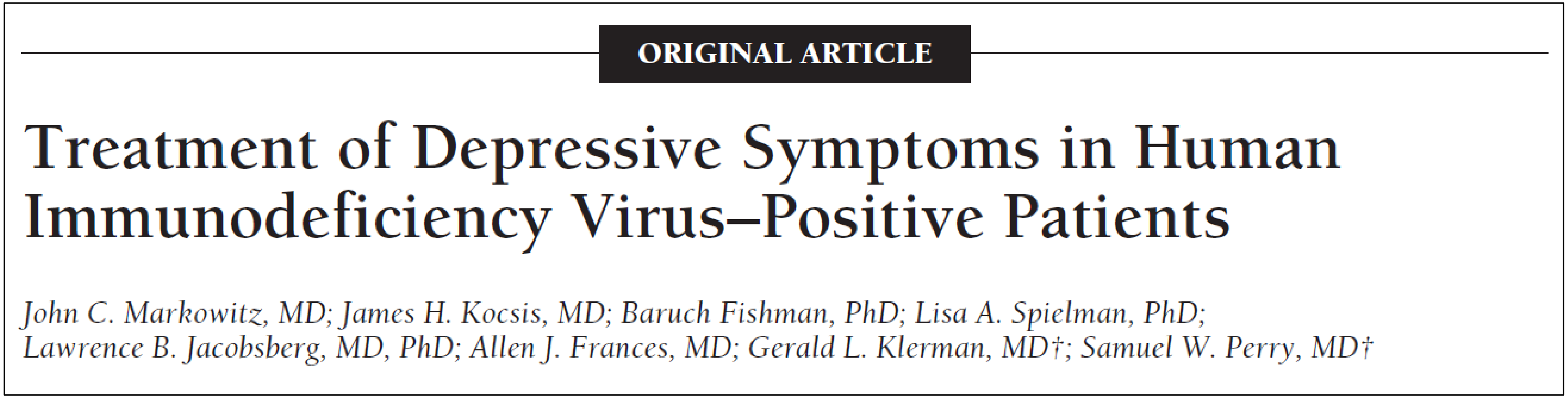
5 February 2020



‘PICO for synthesis’ question:

Is psychotherapy more effective than antidepressants at reducing symptoms of depression?

P	Any person with depressive symptoms
I	Any type of psychotherapy
C	Any type of antidepressant
O	Depression



ORIGINAL ARTICLE

Treatment of Depressive Symptoms in Human Immunodeficiency Virus–Positive Patients

John C. Markowitz, MD; James H. Kocsis, MD; Baruch Fishman, PhD; Lisa A. Spielman, PhD; Lawrence B. Jacobsberg, MD, PhD; Allen J. Frances, MD; Gerald L. Klerman, MD†; Samuel W. Perry, MD†

PICO for 4-arm randomized trial

P People with HIV and depressive symptoms

I One of 3 types of psychotherapy (IPT, CBT, SP)

C Antidepressants (SWI)

O Depression

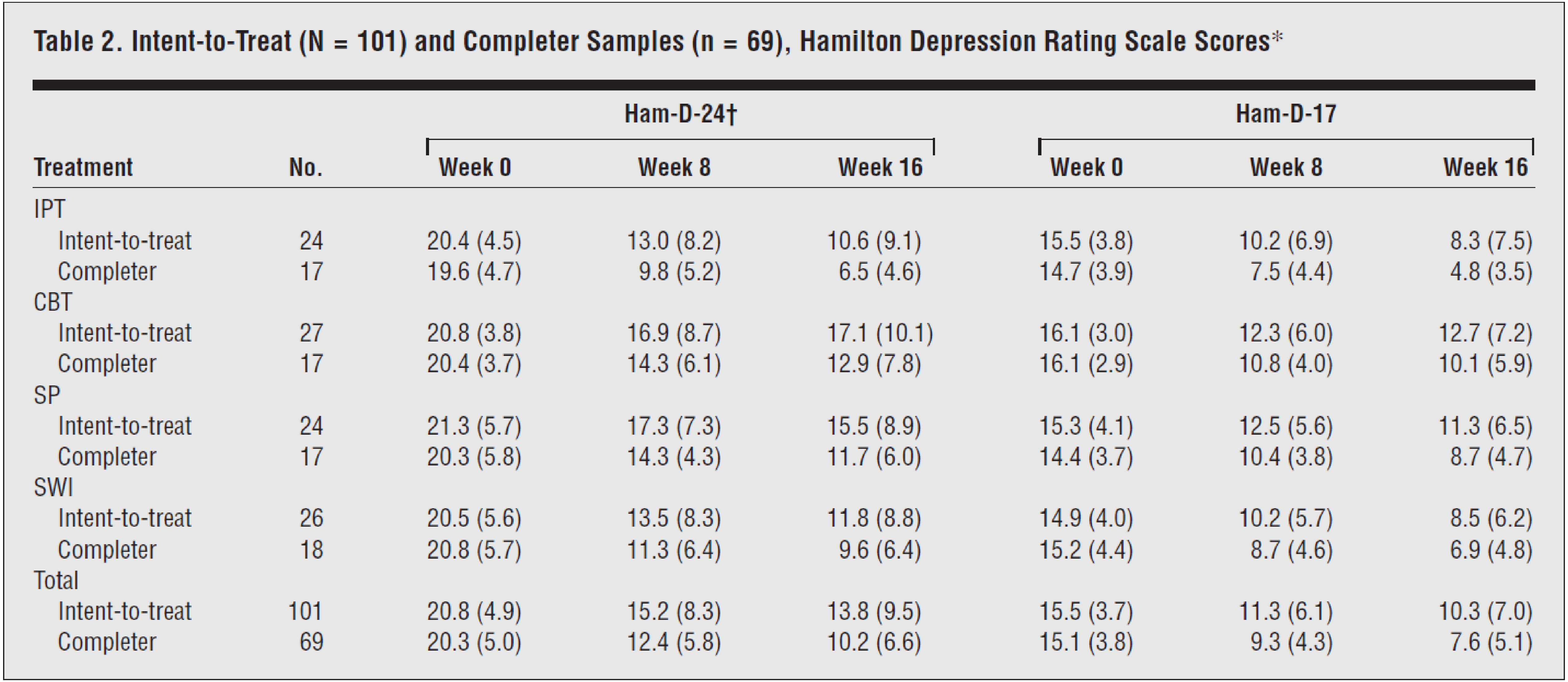


Table 2. Intent-to-Treat (N = 101) and Completer Samples (n = 69), Hamilton Depression Rating Scale Scores*

		Ham-D-24†			Ham-D-17		
Treatment	No.	Week 0	Week 8	Week 16	Week 0	Week 8	Week 16
IPT							
Intent-to-treat	24	20.4 (4.5)	13.0 (8.2)	10.6 (9.1)	15.5 (3.8)	10.2 (6.9)	8.3 (7.5)
Completer	17	19.6 (4.7)	9.8 (5.2)	6.5 (4.6)	14.7 (3.9)	7.5 (4.4)	4.8 (3.5)
CBT							
Intent-to-treat	27	20.8 (3.8)	16.9 (8.7)	17.1 (10.1)	16.1 (3.0)	12.3 (6.0)	12.7 (7.2)
Completer	17	20.4 (3.7)	14.3 (6.1)	12.9 (7.8)	16.1 (2.9)	10.8 (4.0)	10.1 (5.9)
SP							
Intent-to-treat	24	21.3 (5.7)	17.3 (7.3)	15.5 (8.9)	15.3 (4.1)	12.5 (5.6)	11.3 (6.5)
Completer	17	20.3 (5.8)	14.3 (4.3)	11.7 (6.0)	14.4 (3.7)	10.4 (3.8)	8.7 (4.7)
SWI							
Intent-to-treat	26	20.5 (5.6)	13.5 (8.3)	11.8 (8.8)	14.9 (4.0)	10.2 (5.7)	8.5 (6.2)
Completer	18	20.8 (5.7)	11.3 (6.4)	9.6 (6.4)	15.2 (4.4)	8.7 (4.6)	6.9 (4.8)
Total							
Intent-to-treat	101	20.8 (4.9)	15.2 (8.3)	13.8 (9.5)	15.5 (3.7)	11.3 (6.1)	10.3 (7.0)
Completer	69	20.3 (5.0)	12.4 (5.8)	10.2 (6.6)	15.1 (3.8)	9.3 (4.3)	7.6 (5.1)

Markowitz JC, et al. Treatment of depressive symptoms in human immunodeficiency virus-positive patients. Arch Gen Psychiatry 1998;55:452-457.

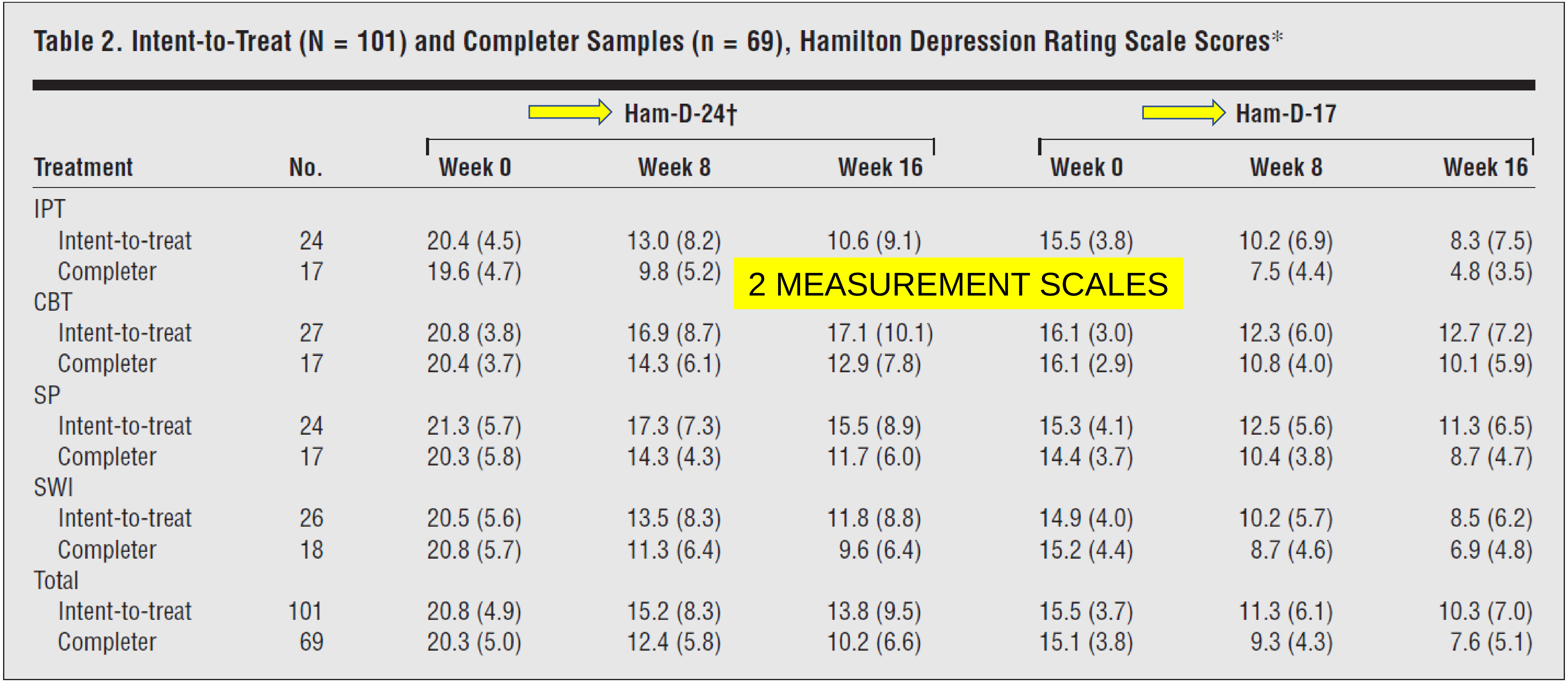


Table 2. Intent-to-Treat (N = 101) and Completer Samples (n = 69), Hamilton Depression Rating Scale Scores*

		Ham-D-24†			Ham-D-17		
Treatment	No.	Week 0	Week 8	Week 16	Week 0	Week 8	Week 16
IPT							
Intent-to-treat	24	20.4 (4.5)	13.0 (8.2)	10.6 (9.1)	15.5 (3.8)	10.2 (6.9)	8.3 (7.5)
Completer	17	19.6 (4.7)	9.8 (5.2)	[illegible]	[illegible]	7.5 (4.4)	4.8 (3.5)
CBT							
Intent-to-treat	27	20.8 (3.8)	16.9 (8.7)	17.1 (10.1)	16.1 (3.0)	12.3 (6.0)	12.7 (7.2)
Completer	17	20.4 (3.7)	14.3 (6.1)	12.9 (7.8)	16.1 (2.9)	10.8 (4.0)	10.1 (5.9)
SP							
Intent-to-treat	24	21.3 (5.7)	17.3 (7.3)	15.5 (8.9)	15.3 (4.1)	12.5 (5.6)	11.3 (6.5)
Completer	17	20.3 (5.8)	14.3 (4.3)	11.7 (6.0)	14.4 (3.7)	10.4 (3.8)	8.7 (4.7)
SWI							
Intent-to-treat	26	20.5 (5.6)	13.5 (8.3)	11.8 (8.8)	14.9 (4.0)	10.2 (5.7)	8.5 (6.2)
Completer	18	20.8 (5.7)	11.3 (6.4)	9.6 (6.4)	15.2 (4.4)	8.7 (4.6)	6.9 (4.8)
Total							
Intent-to-treat	101	20.8 (4.9)	15.2 (8.3)	13.8 (9.5)	15.5 (3.7)	11.3 (6.1)	10.3 (7.0)
Completer	69	20.3 (5.0)	12.4 (5.8)	10.2 (6.6)	15.1 (3.8)	9.3 (4.3)	7.6 (5.1)

2 MEASUREMENT SCALES

Markowitz JC, et al. Treatment of depressive symptoms in human immunodeficiency virus-positive patients. Arch Gen Psychiatry 1998;55:452-457.

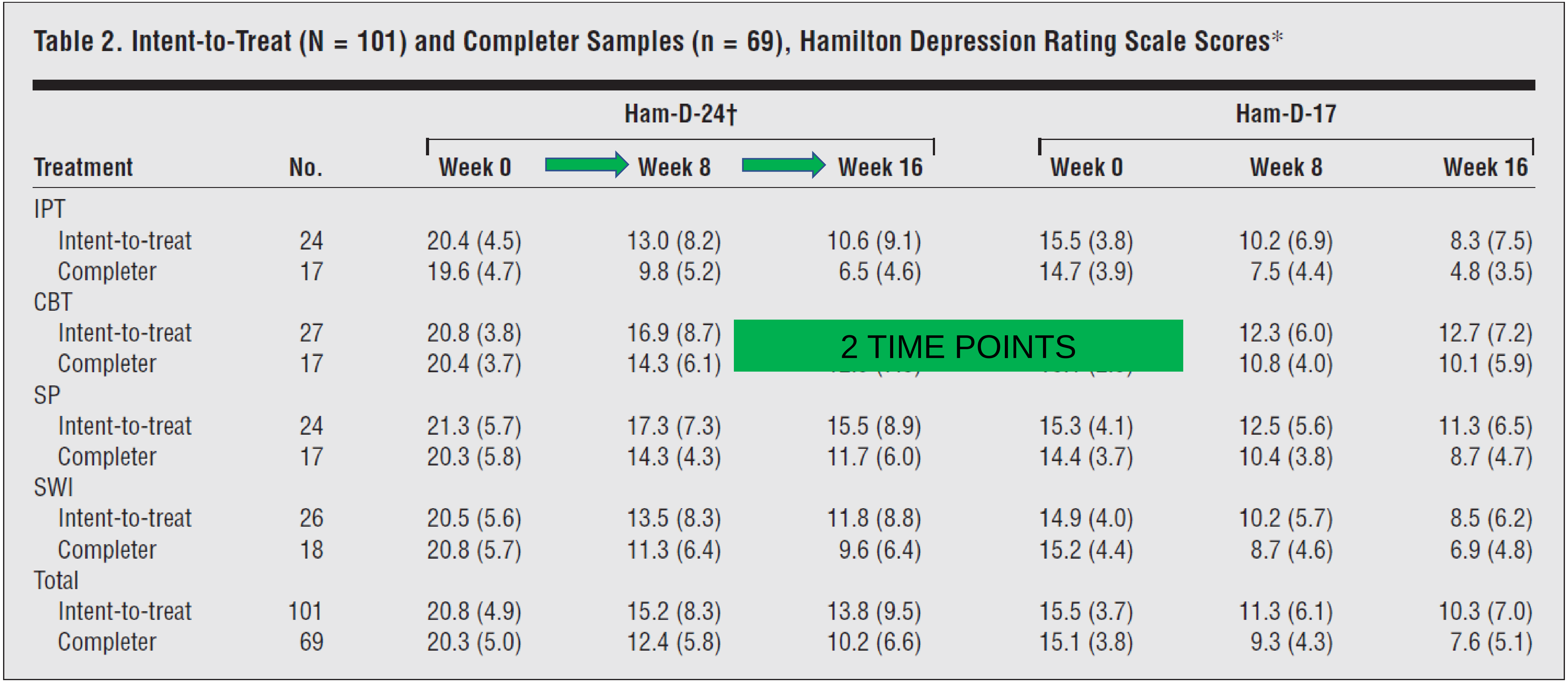


Table 2. Intent-to-Treat (N = 101) and Completer Samples (n = 69), Hamilton Depression Rating Scale Scores*

		Ham-D-24†			Ham-D-17		
Treatment	No.	Week 0	Week 8	Week 16	Week 0	Week 8	Week 16
IPT							
Intent-to-treat	24	20.4 (4.5)	13.0 (8.2)	10.6 (9.1)	15.5 (3.8)	10.2 (6.9)	8.3 (7.5)
Completer	17	19.6 (4.7)	9.8 (5.2)	6.5 (4.6)	14.7 (3.9)	7.5 (4.4)	4.8 (3.5)
CBT							
Intent-to-treat	27	20.8 (3.8)	16.9 (8.7)	[illegible]	[illegible]	12.3 (6.0)	12.7 (7.2)
Completer	17	20.4 (3.7)	14.3 (6.1)	[illegible]	[illegible]	10.8 (4.0)	10.1 (5.9)
SP							
Intent-to-treat	24	21.3 (5.7)	17.3 (7.3)	15.5 (8.9)	15.3 (4.1)	12.5 (5.6)	11.3 (6.5)
Completer	17	20.3 (5.8)	14.3 (4.3)	11.7 (6.0)	14.4 (3.7)	10.4 (3.8)	8.7 (4.7)
SWI							
Intent-to-treat	26	20.5 (5.6)	13.5 (8.3)	11.8 (8.8)	14.9 (4.0)	10.2 (5.7)	8.5 (6.2)
Completer	18	20.8 (5.7)	11.3 (6.4)	9.6 (6.4)	15.2 (4.4)	8.7 (4.6)	6.9 (4.8)
Total							
Intent-to-treat	101	20.8 (4.9)	15.2 (8.3)	13.8 (9.5)	15.5 (3.7)	11.3 (6.1)	10.3 (7.0)
Completer	69	20.3 (5.0)	12.4 (5.8)	10.2 (6.6)	15.1 (3.8)	9.3 (4.3)	7.6 (5.1)

2 TIME POINTS

Markowitz JC, et al. Treatment of depressive symptoms in human immunodeficiency virus-positive patients. Arch Gen Psychiatry 1998;55:452-457.

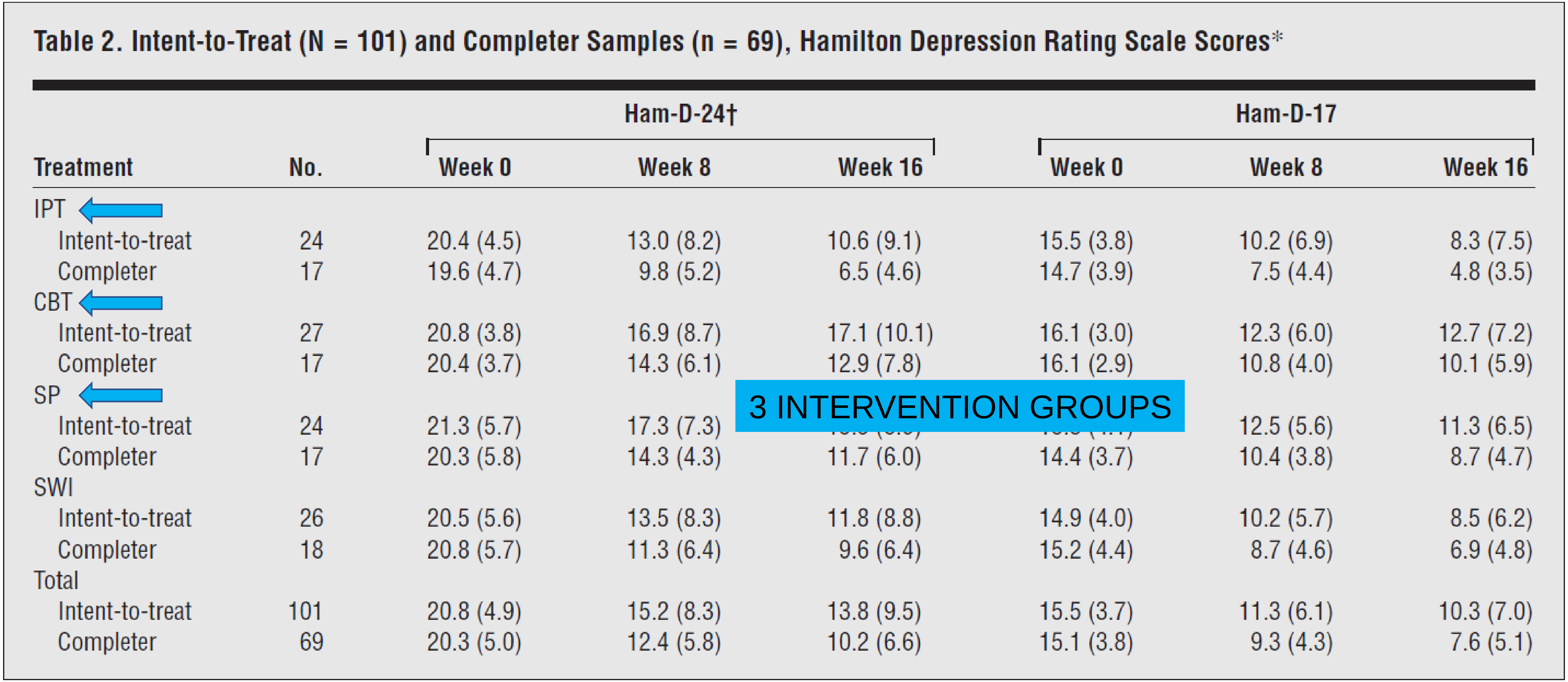


Table 2. Intent-to-Treat (N = 101) and Completer Samples (n = 69), Hamilton Depression Rating Scale Scores*

Treatment	No.	Ham-D-24† Week 0	Ham-D-24† Week 8	Ham-D-24† Week 16	Ham-D-17 Week 0	Ham-D-17 Week 8	Ham-D-17 Week 16
IPT							
Intent-to-treat	24	20.4 (4.5)	13.0 (8.2)	10.6 (9.1)	15.5 (3.8)	10.2 (6.9)	8.3 (7.5)
Completer	17	19.6 (4.7)	9.8 (5.2)	6.5 (4.6)	14.7 (3.9)	7.5 (4.4)	4.8 (3.5)
CBT							
Intent-to-treat	27	20.8 (3.8)	16.9 (8.7)	17.1 (10.1)	16.1 (3.0)	12.3 (6.0)	12.7 (7.2)
Completer	17	20.4 (3.7)	14.3 (6.1)	12.9 (7.8)	16.1 (2.9)	10.8 (4.0)	10.1 (5.9)
SP							
Intent-to-treat	24	21.3 (5.7)	17.3 (7.3)	[illegible]	[illegible]	12.5 (5.6)	11.3 (6.5)
Completer	17	20.3 (5.8)	14.3 (4.3)	11.7 (6.0)	14.4 (3.7)	10.4 (3.8)	8.7 (4.7)
SWI							
Intent-to-treat	26	20.5 (5.6)	13.5 (8.3)	11.8 (8.8)	14.9 (4.0)	10.2 (5.7)	8.5 (6.2)
Completer	18	20.8 (5.7)	11.3 (6.4)	9.6 (6.4)	15.2 (4.4)	8.7 (4.6)	6.9 (4.8)
Total							
Intent-to-treat	101	20.8 (4.9)	15.2 (8.3)	13.8 (9.5)	15.5 (3.7)	11.3 (6.1)	10.3 (7.0)
Completer	69	20.3 (5.0)	12.4 (5.8)	10.2 (6.6)	15.1 (3.8)	9.3 (4.3)	7.6 (5.1)

3 INTERVENTION GROUPS

Markowitz JC, et al. Treatment of depressive symptoms in human immunodeficiency virus-positive patients. Arch Gen Psychiatry 1998;55:452-457.

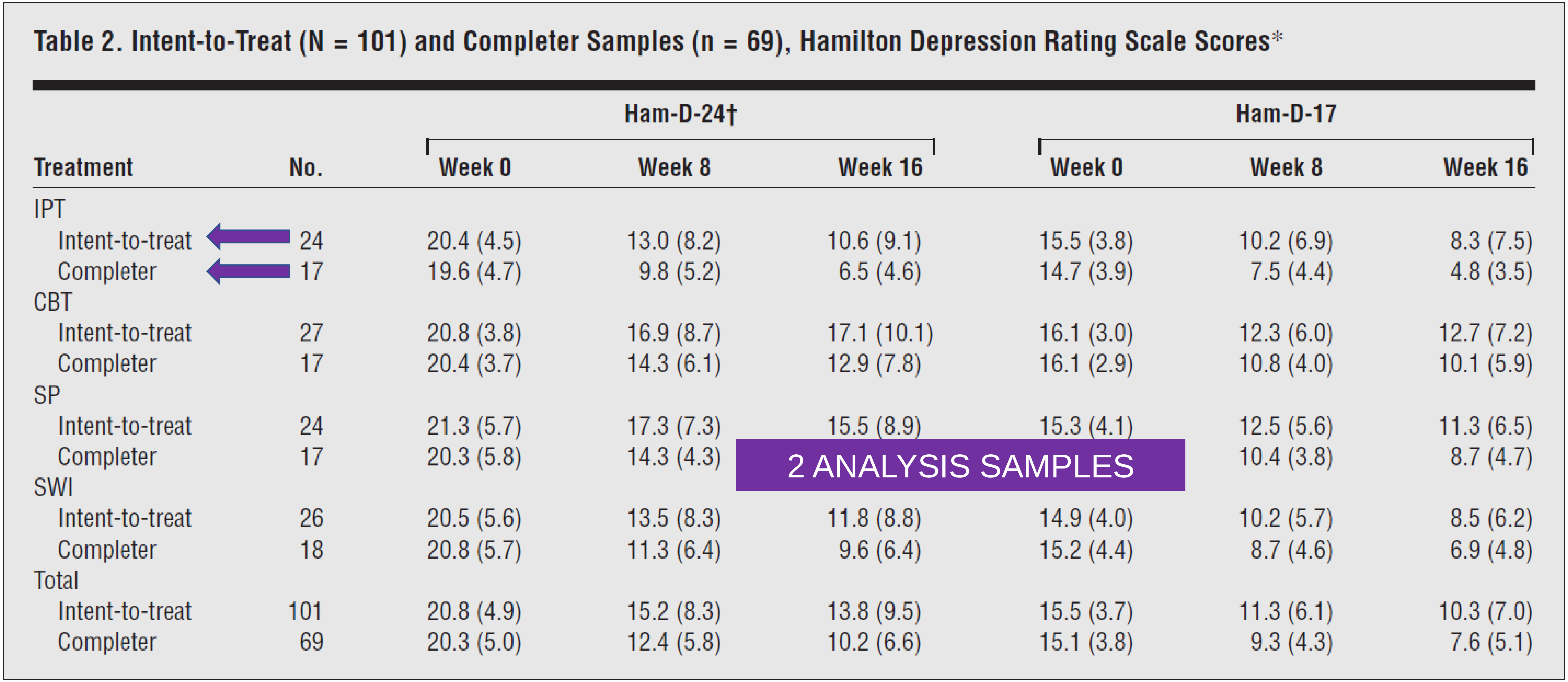


Table 2. Intent-to-Treat (N = 101) and Completer Samples (n = 69), Hamilton Depression Rating Scale Scores*

		Ham-D-24†			Ham-D-17		
Treatment	No.	Week 0	Week 8	Week 16	Week 0	Week 8	Week 16
IPT							
Intent-to-treat	24	20.4 (4.5)	13.0 (8.2)	10.6 (9.1)	15.5 (3.8)	10.2 (6.9)	8.3 (7.5)
Completer	17	19.6 (4.7)	9.8 (5.2)	6.5 (4.6)	14.7 (3.9)	7.5 (4.4)	4.8 (3.5)
CBT							
Intent-to-treat	27	20.8 (3.8)	16.9 (8.7)	17.1 (10.1)	16.1 (3.0)	12.3 (6.0)	12.7 (7.2)
Completer	17	20.4 (3.7)	14.3 (6.1)	12.9 (7.8)	16.1 (2.9)	10.8 (4.0)	10.1 (5.9)
SP							
Intent-to-treat	24	21.3 (5.7)	17.3 (7.3)	15.5 (8.9)	15.3 (4.1)	12.5 (5.6)	11.3 (6.5)
Completer	17	20.3 (5.8)	14.3 (4.3)	[illegible]	[illegible]	10.4 (3.8)	8.7 (4.7)
SWI							
Intent-to-treat	26	20.5 (5.6)	13.5 (8.3)	11.8 (8.8)	14.9 (4.0)	10.2 (5.7)	8.5 (6.2)
Completer	18	20.8 (5.7)	11.3 (6.4)	9.6 (6.4)	15.2 (4.4)	8.7 (4.6)	6.9 (4.8)
Total							
Intent-to-treat	101	20.8 (4.9)	15.2 (8.3)	13.8 (9.5)	15.5 (3.7)	11.3 (6.1)	10.3 (7.0)
Completer	69	20.3 (5.0)	12.4 (5.8)	10.2 (6.6)	15.1 (3.8)	9.3 (4.3)	7.6 (5.1)

2 ANALYSIS SAMPLES

Markowitz JC, et al. Treatment of depressive symptoms in human immunodeficiency virus-positive patients. Arch Gen Psychiatry 1998;55:452-457.

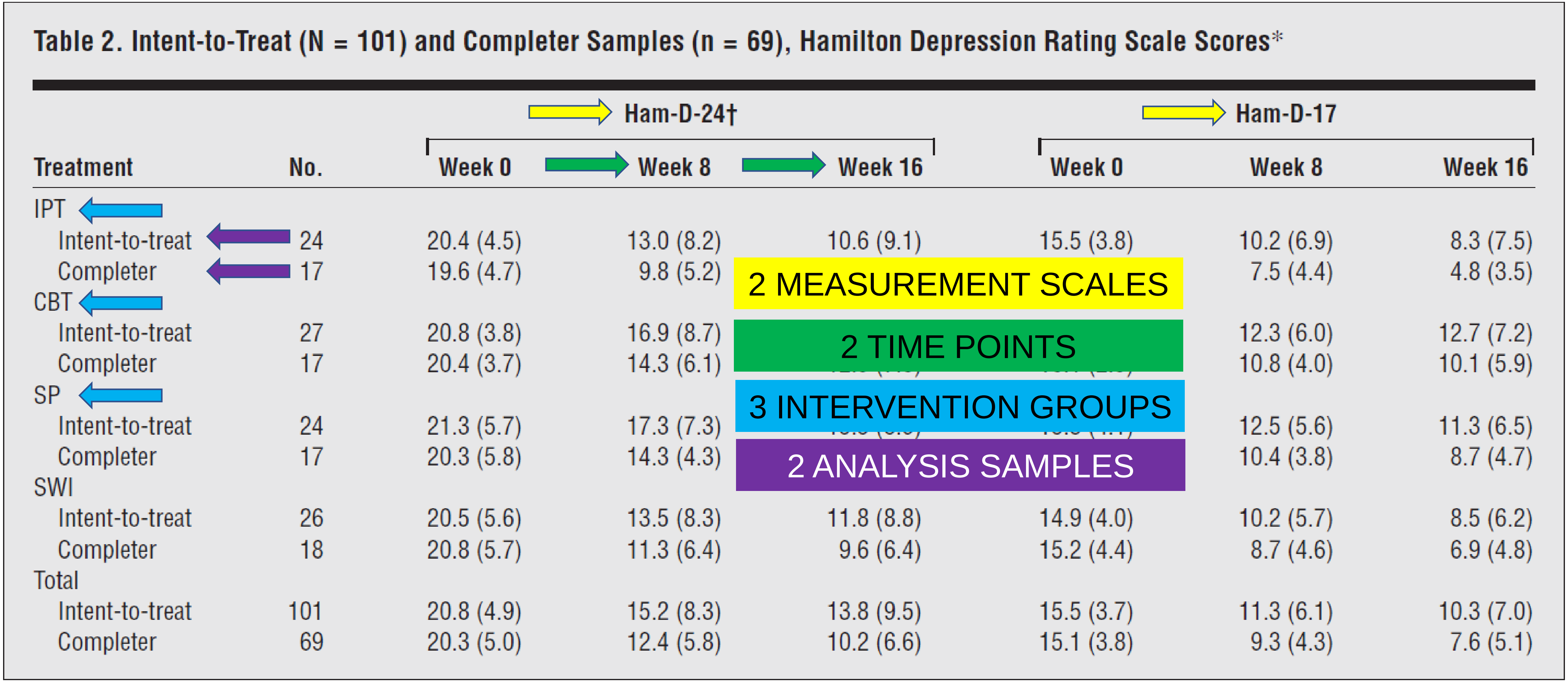


Table 2. Intent-to-Treat (N = 101) and Completer Samples (n = 69), Hamilton Depression Rating Scale Scores*

Treatment	No.	Ham-D-24†			Ham-D-17		
		Week 0	Week 8	Week 16	Week 0	Week 8	Week 16
IPT							
Intent-to-treat	24	20.4 (4.5)	13.0 (8.2)	10.6 (9.1)	15.5 (3.8)	10.2 (6.9)	8.3 (7.5)
Completer	17	19.6 (4.7)	9.8 (5.2)	[illegible]	[illegible]	7.5 (4.4)	4.8 (3.5)
CBT							
Intent-to-treat	27	20.8 (3.8)	16.9 (8.7)	[illegible]	[illegible]	12.3 (6.0)	12.7 (7.2)
Completer	17	20.4 (3.7)	14.3 (6.1)	[illegible]	[illegible]	10.8 (4.0)	10.1 (5.9)
SP							
Intent-to-treat	24	21.3 (5.7)	17.3 (7.3)	[illegible]	[illegible]	12.5 (5.6)	11.3 (6.5)
Completer	17	20.3 (5.8)	14.3 (4.3)	[illegible]	[illegible]	10.4 (3.8)	8.7 (4.7)
SWI							
Intent-to-treat	26	20.5 (5.6)	13.5 (8.3)	11.8 (8.8)	14.9 (4.0)	10.2 (5.7)	8.5 (6.2)
Completer	18	20.8 (5.7)	11.3 (6.4)	9.6 (6.4)	15.2 (4.4)	8.7 (4.6)	6.9 (4.8)
Total							
Intent-to-treat	101	20.8 (4.9)	15.2 (8.3)	13.8 (9.5)	15.5 (3.7)	11.3 (6.1)	10.3 (7.0)
Completer	69	20.3 (5.0)	12.4 (5.8)	10.2 (6.6)	15.1 (3.8)	9.3 (4.3)	7.6 (5.1)

2 MEASUREMENT SCALES

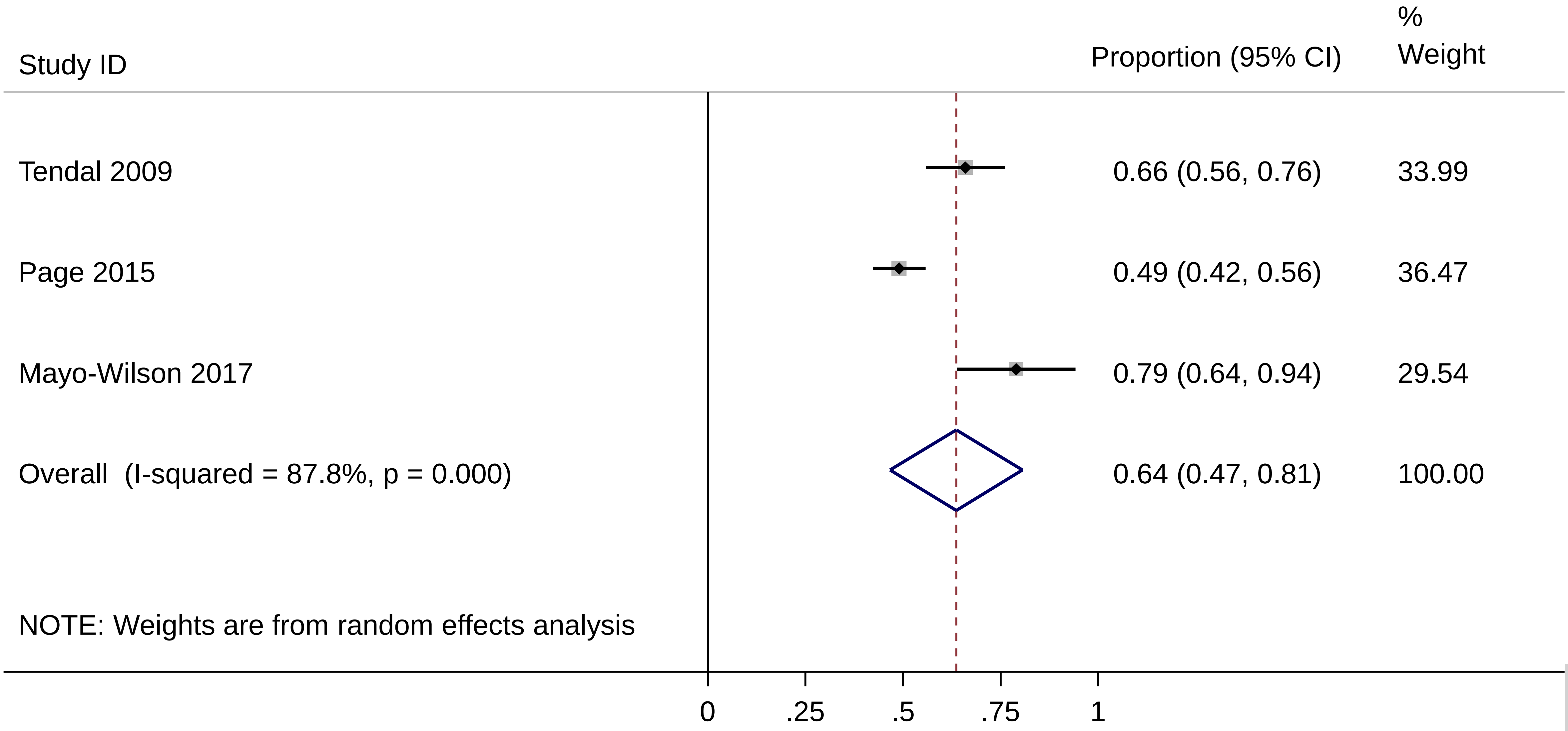
2 TIME POINTS

3 INTERVENTION GROUPS

2 ANALYSIS SAMPLES

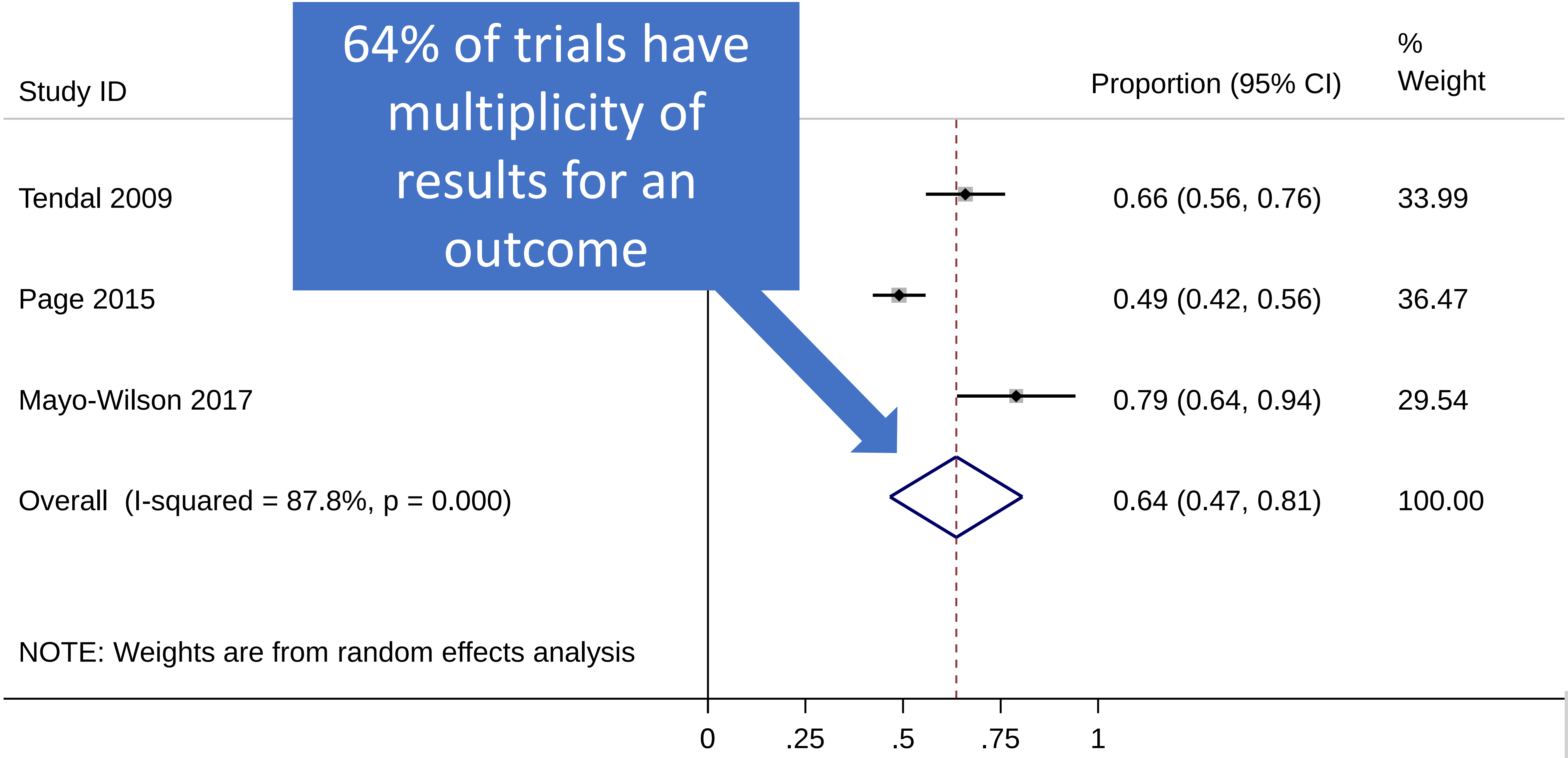
Markowitz JC, et al. Treatment of depressive symptoms in human immunodeficiency virus-positive patients. Arch Gen Psychiatry 1998;55:452-457.

Trial has >24 effect estimates for depression!



Study ID	Proportion (95% CI)	% Weight
Tendal 2009	0.66 (0.56, 0.76)	33.99
Page 2015	0.49 (0.42, 0.56)	36.47
Mayo-Wilson 2017	0.79 (0.64, 0.94)	29.54
Overall (I-squared = 87.8%, p = 0.000)	0.64 (0.47, 0.81)	100.00

NOTE: Weights are from random effects analysis



64% of trials have multiplicity of results for an outcome

Study ID	Proportion (95% CI)	% Weight
Tendal 2009	0.66 (0.56, 0.76)	33.99
Page 2015	0.49 (0.42, 0.56)	36.47
Mayo-Wilson 2017	0.79 (0.64, 0.94)	29.54
Overall (I-squared = 87.8%, p = 0.000)	0.64 (0.47, 0.81)	100.00

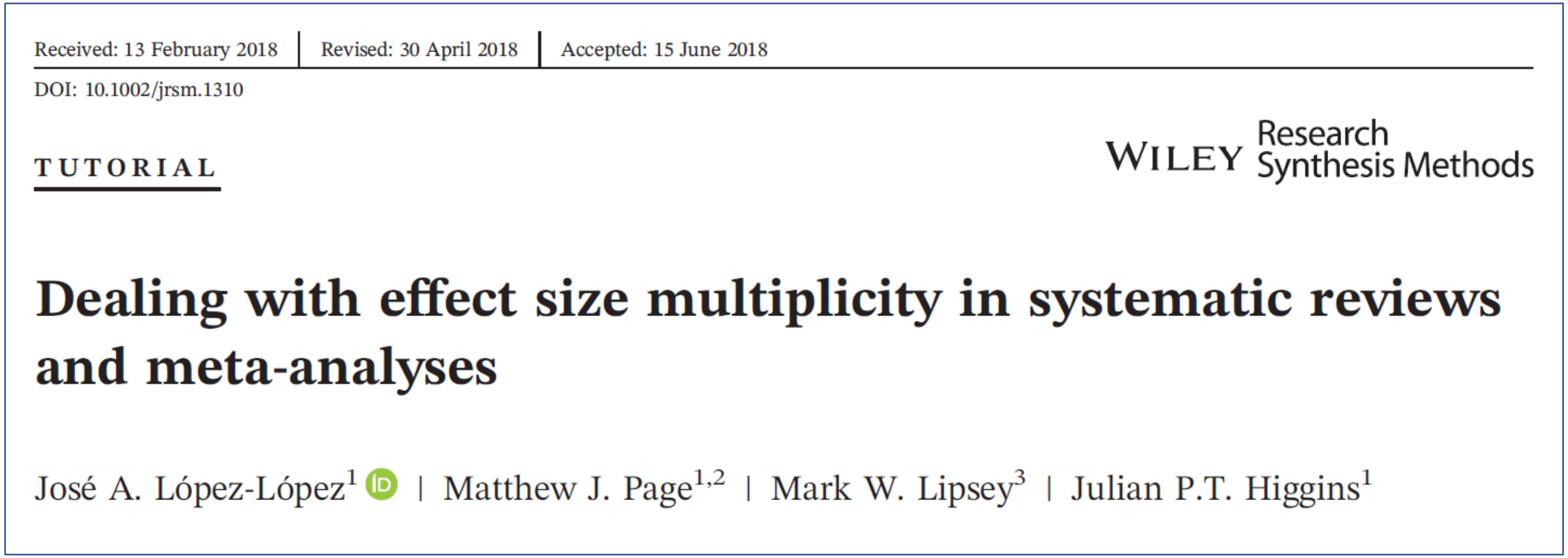
NOTE: Weights are from random effects analysis

Why this matters

Multiplicity can lead to **selective inclusion of results**, where systematic reviewers' choice about which result to include are influenced by the P value, magnitude or direction of the results

Can also lead to inconsistencies between reviewers in the data collected (in the absence of bias)

Page MJ et al. Investigation of bias in meta-analyses due to selective inclusion of trial effect estimates: empirical study. BMJ Open 2016;6(4):e011863



Received: 13 February 2018 | Revised: 30 April 2018 | Accepted: 15 June 2018

DOI: 10.1002/jrsm.1310

TUTORIAL

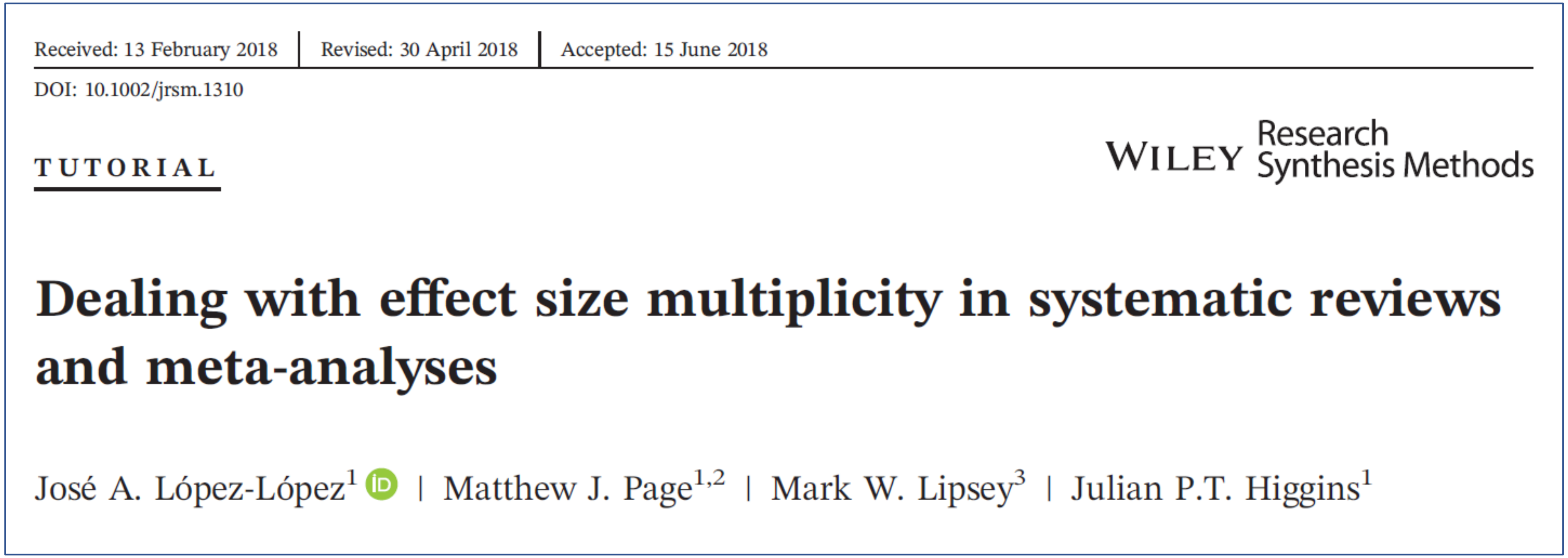
WILEY Research Synthesis Methods

Dealing with effect size multiplicity in systematic reviews and meta-analyses

José A. López-López[1] | Matthew J. Page[1,2] | Mark W. Lipsey[3] | Julian P.T. Higgins[1]

Two meta-analytic approaches for dealing with multiplicity

1. Reductionist approach: inclusion of a single effect estimate per study
2. Integrative approach: inclusion of multiple effect estimates per study



Received: 13 February 2018 | Revised: 30 April 2018 | Accepted: 15 June 2018

DOI: 10.1002/jrsm.1310

TUTORIAL

WILEY Research Synthesis Methods

Dealing with effect size multiplicity in systematic reviews and meta-analyses

José A. López-López[1] | Matthew J. Page[1,2] | Mark W. Lipsey[3] | Julian P.T. Higgins[1]

Two meta-analytic approaches for dealing with multiplicity

1. Reductionist approach: inclusion of a single effect estimate per study
2. Integrative approach: inclusion of multiple effect estimates per study

Hierarchical selection rules

Pre-defined strategies to select one effect estimate from a study when multiple estimates are encountered

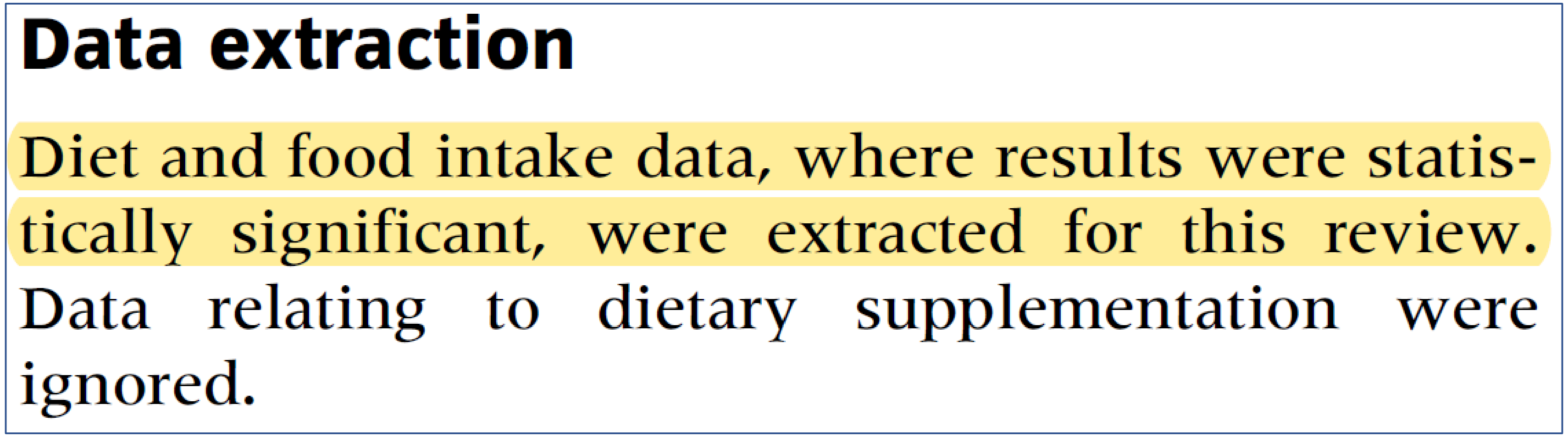
Appropriate when multiple effect estimates are regarded as being loosely equivalent but not completely interchangeable

Rules should be based on clinical or methodological rationale

- e.g. plan to select measurement scales with the best measurement properties, at time points that are most clinically relevant

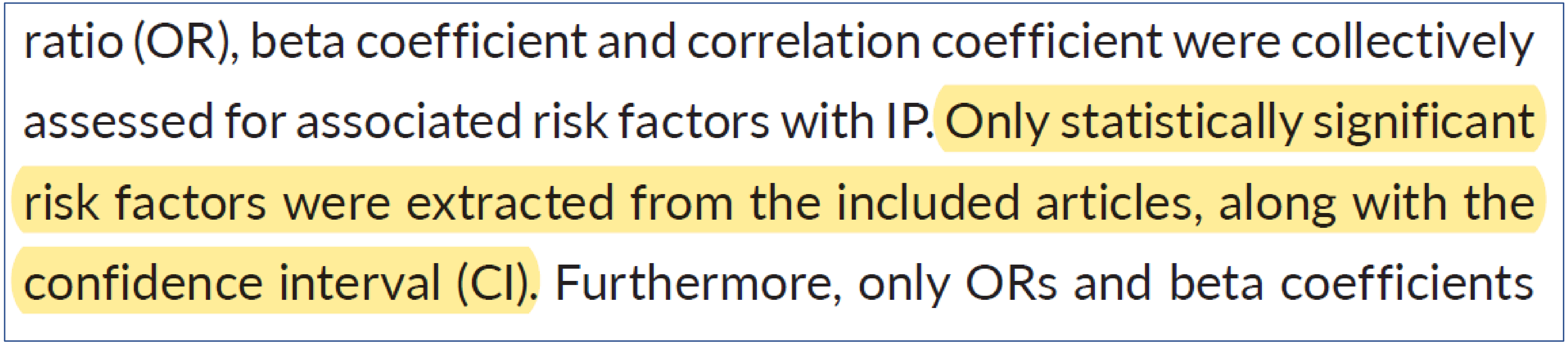
López-López JA, et al. Dealing with effect size multiplicity in systematic reviews and meta-analyses. Res Synth Methods 2018. doi: 10.1002/jrsm.1310

Unacceptable selection rules



Data extraction

Diet and food intake data, where results were statistically significant, were extracted for this review. Data relating to dietary supplementation were ignored.



ratio (OR), beta coefficient and correlation coefficient were collectively assessed for associated risk factors with IP. Only statistically significant risk factors were extracted from the included articles, along with the confidence interval (CI). Furthermore, only ORs and beta coefficients

Acceptable selection rules

“Where trialists reported outcome data for more than one function scale, we extracted data on the scale that was highest on the following *a priori* defined list:

- Shoulder Pain and Disability Index (SPADI)
- Croft Shoulder Disability Questionnaire
- Constant-Murley Score
- Any other shoulder-specific function scale”

“If data were available in a trial at multiple time points within each of the above periods (e.g. at four, five, and six weeks), we only extracted data at the latest possible time point of each period”

Page MJ et al. Manual therapy and exercise for rotator cuff disease.
Cochrane Database of Systematic Reviews 2016, Issue 6. Art.No.:CD012224.

Acceptable selection rules

“Where trialists reported outcome data for more than one function scale, we extracted data on the scale that was highest on the following *a priori* defined list:

- Shoulder Pain and Disability Index (SPADI)
- Croft Shoulder Disability Questionnaire
- Constant-Murley Score
- Any other shoulder-specific function scale”

Based on measurement properties

“If data were available in a trial at multiple time points within each of the above periods (e.g. at four, five, and six weeks), we only extracted data at the latest possible time point of each period”

Page MJ et al. Manual therapy and exercise for rotator cuff disease.
Cochrane Database of Systematic Reviews 2016, Issue 6. Art.No.:CD012224.

Acceptable selection rules

“Where trialists reported outcome data for more than one function scale, we extracted data on the scale that was highest on the following *a priori* defined list:

- Shoulder Pain and Disability Index (SPADI)
- Croft Shoulder Disability Questionnaire
- Constant-Murley Score
- Any other shoulder-specific function scale”

“If data were available in a trial at multiple time points within each of the above periods (e.g. at four, five, and six weeks), we only extracted data at the latest possible time point of each period”

Page MJ et al. Manual therapy and exercise for rotator cuff disease.
Cochrane Database of Systematic Reviews 2016, Issue 6. Art.No.:CD012224.

Acceptable selection rules

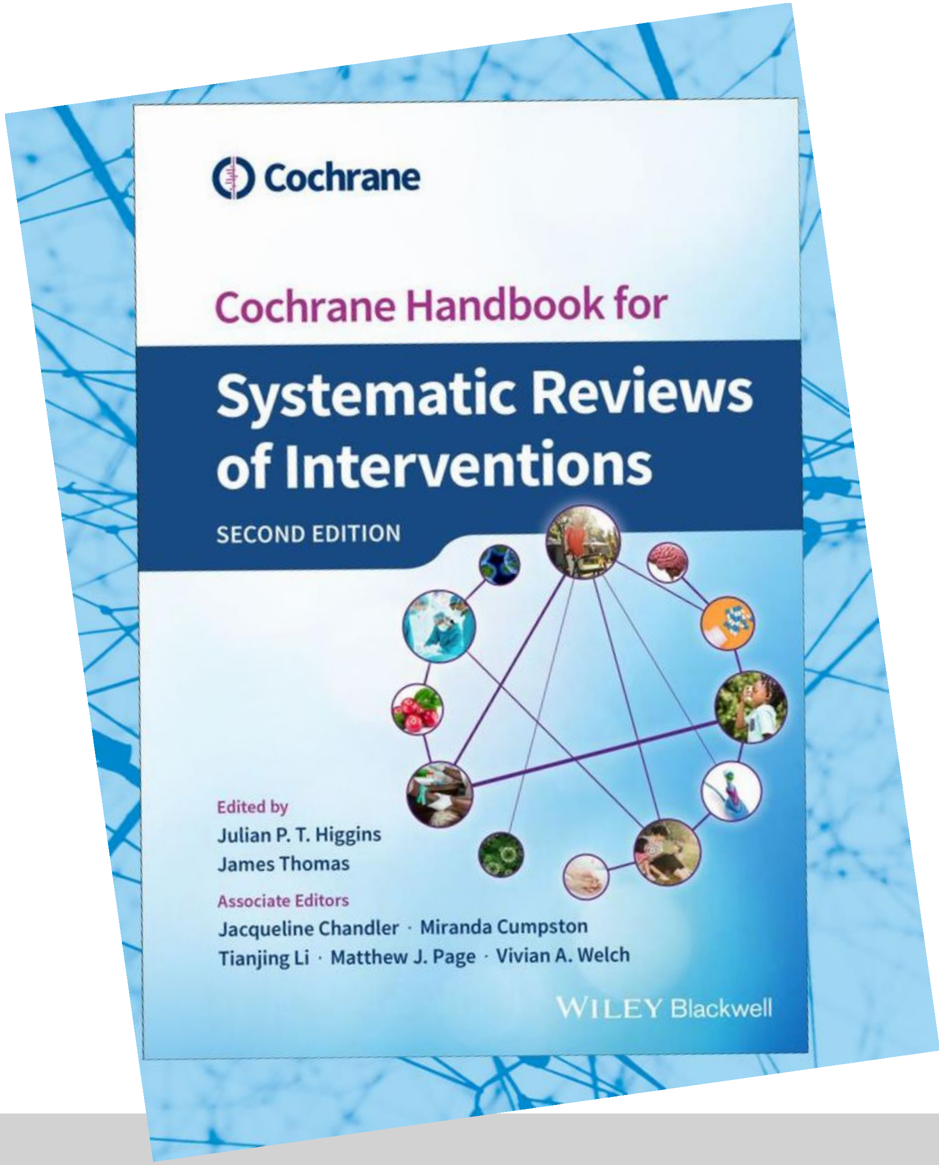
"Where trialists reported outcome data for more than one function scale, we extracted data on the scale that was highest on the following *a priori* defined list:

- Shoulder Pain and Disability Index (SPADI)
- Croft Shoulder Disability Questionnaire
- Constant-Murley Score
- Any other shoulder-specific function scale"

Considered most clinically important time point

"If data were available in a trial at multiple time points within each of the above periods (e.g. at four, five, and six weeks), we only extracted data at the latest possible time point of each period"

Page MJ et al. Manual therapy and exercise for rotator cuff disease.
Cochrane Database of Systematic Reviews 2016, Issue 6. Art.No.:CD012224.



Cochrane

Cochrane Handbook for

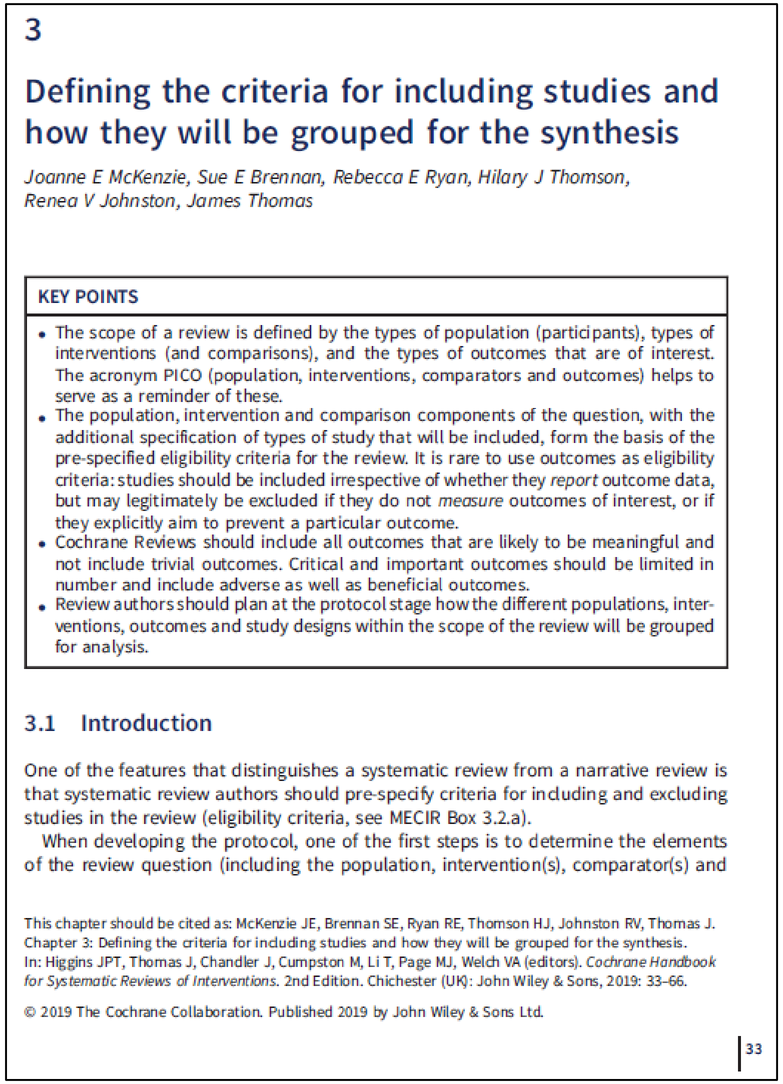
Systematic Reviews of Interventions

SECOND EDITION

Edited by
Julian P. T. Higgins
James Thomas

Associate Editors
Jacqueline Chandler · Miranda Cumpston
Tianjing Li · Matthew J. Page · Vivian A. Welch

WILEY Blackwell



3

Defining the criteria for including studies and how they will be grouped for the synthesis

Joanne E McKenzie, Sue E Brennan, Rebecca E Ryan, Hilary J Thomson, Renea V Johnston, James Thomas

KEY POINTS

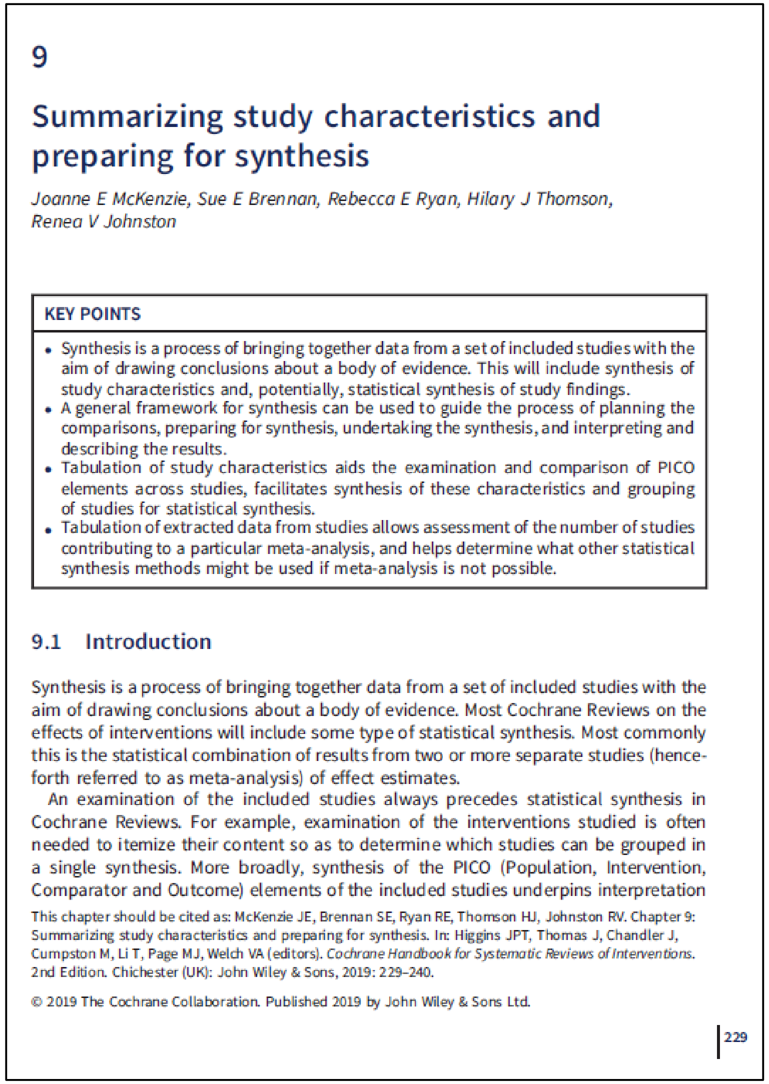
- The scope of a review is defined by the types of population (participants), types of interventions (and comparisons), and the types of outcomes that are of interest. The acronym PICO (population, interventions, comparators and outcomes) helps to serve as a reminder of these.
- The population, intervention and comparison components of the question, with the additional specification of types of study that will be included, form the basis of the pre-specified eligibility criteria for the review. It is rare to use outcomes as eligibility criteria: studies should be included irrespective of whether they *report* outcome data, but may legitimately be excluded if they do not *measure* outcomes of interest, or if they explicitly aim to prevent a particular outcome.
- Cochrane Reviews should include all outcomes that are likely to be meaningful and not include trivial outcomes. Critical and important outcomes should be limited in number and include adverse as well as beneficial outcomes.
- Review authors should plan at the protocol stage how the different populations, interventions, outcomes and study designs within the scope of the review will be grouped for analysis.

3.1 Introduction

One of the features that distinguishes a systematic review from a narrative review is that systematic review authors should pre-specify criteria for including and excluding studies in the review (eligibility criteria, see MECIR Box 3.2.a).

When developing the protocol, one of the first steps is to determine the elements of the review question (including the population, intervention(s), comparator(s) and

This chapter should be cited as: McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV, Thomas J. Chapter 3: Defining the criteria for including studies and how they will be grouped for the synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019: 33–66.



9

Summarizing study characteristics and preparing for synthesis

Joanne E McKenzie, Sue E Brennan, Rebecca E Ryan, Hilary J Thomson, Renea V Johnston

KEY POINTS

- Synthesis is a process of bringing together data from a set of included studies with the aim of drawing conclusions about a body of evidence. This will include synthesis of study characteristics and, potentially, statistical synthesis of study findings.
- A general framework for synthesis can be used to guide the process of planning the comparisons, preparing for synthesis, undertaking the synthesis, and interpreting and describing the results.
- Tabulation of study characteristics aids the examination and comparison of PICO elements across studies, facilitates synthesis of these characteristics and grouping of studies for statistical synthesis.
- Tabulation of extracted data from studies allows assessment of the number of studies contributing to a particular meta-analysis, and helps determine what other statistical synthesis methods might be used if meta-analysis is not possible.

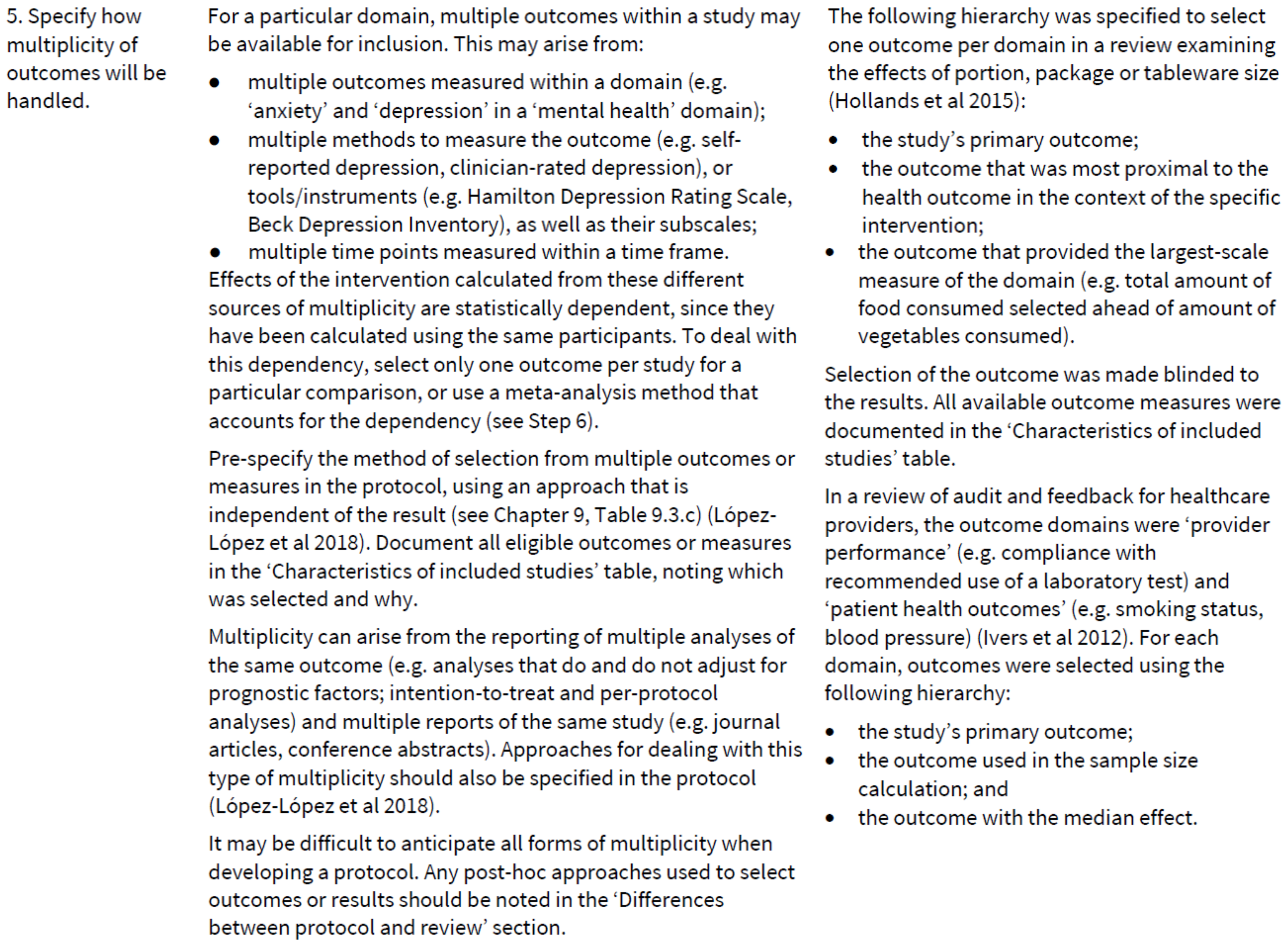
9.1 Introduction

Synthesis is a process of bringing together data from a set of included studies with the aim of drawing conclusions about a body of evidence. Most Cochrane Reviews on the effects of interventions will include some type of statistical synthesis. Most commonly this is the statistical combination of results from two or more separate studies (henceforth referred to as meta-analysis) of effect estimates.

An examination of the included studies always precedes statistical synthesis in Cochrane Reviews. For example, examination of the interventions studied is often needed to itemize their content so as to determine which studies can be grouped in a single synthesis. More broadly, synthesis of the PICO (Population, Intervention, Comparator and Outcome) elements of the included studies underpins interpretation

This chapter should be cited as: McKenzie JE, Brennan SE, Ryan RE, Thomson HJ, Johnston RV. Chapter 9: Summarizing study characteristics and preparing for synthesis. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019: 229–240.

1. Fully specify outcome domains
2. Determine whether there is an existing system for identifying and grouping important outcomes
3. Define the outcome time points
4. Specify the measurement tool or measurement method
5. Specify how multiplicity of outcomes will be handled
6. Plan how the specified outcome domains will be used in the synthesis
7. Build in contingencies by specifying both specific and broader outcome domains



5. Specify how multiplicity of outcomes will be handled.	For a particular domain, multiple outcomes within a study may be available for inclusion. This may arise from: • multiple outcomes measured within a domain (e.g. 'anxiety' and 'depression' in a 'mental health' domain); • multiple methods to measure the outcome (e.g. self-reported depression, clinician-rated depression), or tools/instruments (e.g. Hamilton Depression Rating Scale, Beck Depression Inventory), as well as their subscales; • multiple time points measured within a time frame. Effects of the intervention calculated from these different sources of multiplicity are statistically dependent, since they have been calculated using the same participants. To deal with this dependency, select only one outcome per study for a particular comparison, or use a meta-analysis method that accounts for the dependency (see Step 6). Pre-specify the method of selection from multiple outcomes or measures in the protocol, using an approach that is independent of the result (see Chapter 9, Table 9.3.c) (López-López et al 2018). Document all eligible outcomes or measures in the 'Characteristics of included studies' table, noting which was selected and why. Multiplicity can arise from the reporting of multiple analyses of the same outcome (e.g. analyses that do and do not adjust for prognostic factors; intention-to-treat and per-protocol analyses) and multiple reports of the same study (e.g. journal articles, conference abstracts). Approaches for dealing with this type of multiplicity should also be specified in the protocol (López-López et al 2018). It may be difficult to anticipate all forms of multiplicity when developing a protocol. Any post-hoc approaches used to select outcomes or results should be noted in the 'Differences between protocol and review' section.	The following hierarchy was specified to select one outcome per domain in a review examining the effects of portion, package or tableware size (Hollands et al 2015): • the study's primary outcome; • the outcome that was most proximal to the health outcome in the context of the specific intervention; • the outcome that provided the largest-scale measure of the domain (e.g. total amount of food consumed selected ahead of amount of vegetables consumed). Selection of the outcome was made blinded to the results. All available outcome measures were documented in the 'Characteristics of included studies' table. In a review of audit and feedback for healthcare providers, the outcome domains were 'provider performance' (e.g. compliance with recommended use of a laboratory test) and 'patient health outcomes' (e.g. smoking status, blood pressure) (Ivers et al 2012). For each domain, outcomes were selected using the following hierarchy: • the study's primary outcome; • the outcome used in the sample size calculation; and • the outcome with the median effect.

Cochrane Handbook Version 6, Chapter 3, Table 3.2.c
https://training.cochrane.org/handbook/current/chapter-03

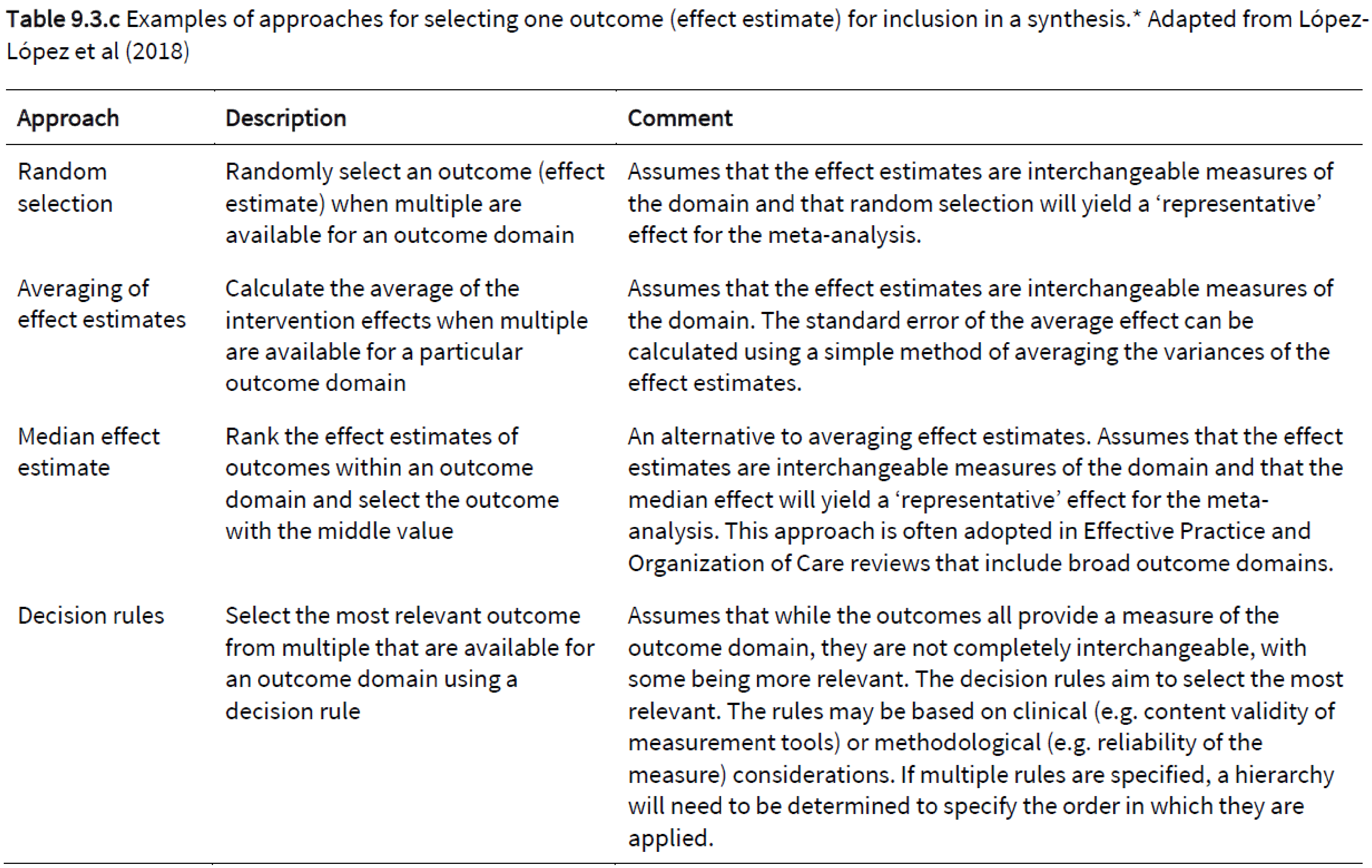


Table 9.3.c Examples of approaches for selecting one outcome (effect estimate) for inclusion in a synthesis.* Adapted from López-López et al (2018)

Approach	Description	Comment
Random selection	Randomly select an outcome (effect estimate) when multiple are available for an outcome domain	Assumes that the effect estimates are interchangeable measures of the domain and that random selection will yield a 'representative' effect for the meta-analysis.
Averaging of effect estimates	Calculate the average of the intervention effects when multiple are available for a particular outcome domain	Assumes that the effect estimates are interchangeable measures of the domain. The standard error of the average effect can be calculated using a simple method of averaging the variances of the effect estimates.
Median effect estimate	Rank the effect estimates of outcomes within an outcome domain and select the outcome with the middle value	An alternative to averaging effect estimates. Assumes that the effect estimates are interchangeable measures of the domain and that the median effect will yield a 'representative' effect for the meta-analysis. This approach is often adopted in Effective Practice and Organization of Care reviews that include broad outcome domains.
Decision rules	Select the most relevant outcome from multiple that are available for an outcome domain using a decision rule	Assumes that while the outcomes all provide a measure of the outcome domain, they are not completely interchangeable, with some being more relevant. The decision rules aim to select the most relevant. The rules may be based on clinical (e.g. content validity of measurement tools) or methodological (e.g. reliability of the measure) considerations. If multiple rules are specified, a hierarchy will need to be determined to specify the order in which they are applied.

Cochrane Handbook Version 6, Chapter 9, Table 9.3.c
https://training.cochrane.org/handbook/current/chapter-09

Conclusion

Hierarchical selection rules can reduce risk of bias in meta-analyses due to selective inclusion of results

May save a lot of time

- No need to extract data for all results in studies
- Less time needed to sort data for synthesis

Might need to revise plans if rules do not suit the data observed; ensure post-hoc rules do not systematically select estimates based on P value, magnitude or direction of effect